SALIENT OBJECT DETECTION FROM THE VIDEO STREAMS USING WATER FLOW APPROACH

A Synopsis

Submitted in

partial fulfillment for the degree

of

Doctor of Philosophy

(Computer Science)

Supervised by

JV'n Dr. Kavita

Submitted by

JV'n Ms. Ruchi Kshatri

Department of Computer Science

Faculty of Engineering and Technology

Jayoti Vidyapeeth Women's University, Jaipur (Rajasthan)

August , 2018

1. Introduction

The recent development in the computer vision application is to detect the salient object in the video streams which plays a vital role nowadays. When human seeing an image, they will focus on the important and informative region of an image. In order to focus on the foreground object without any background deviation, we are going to the Salient Object Detection (SOD). The major task of SOD is to find an important part of the image from the background priors which takes place at the time of the image preprocessing such as image segmentation, visual tracking etc., SOD have two kind of categories, namely, Top- Down and Bottom-Up approach. The Top-Down is related to the supervised learning whereas the Bottom-Up relies on colors, features and spatial distance to build the salient map [1].

The SOD model lies on up-to-date applications like computer vision, Robotics and graphic fields where the phenomenon of identifying the attentive region in the static/dynamic video or image are said to be Salient region detection [2]. It will usually afford to generate the smooth connected areas. To evaluate this region, the first one is to find the large salient area in the frame which results to the false positive for the fixation prediction. The second one is that when popping-out is done only the sparse salient region will be missing in the detection of salient regions of objects. Visual saliency is applied into various applications such as Stylish rendering, Object detection and Image segmentation [3].

Although image boundary is commonly used as background priors, it does not work well for images of complex scenes and video, and it relies a challenging task. Here most offered visual attention approaches are based on the bottom-up computational framework. The idea of bottom-up framework is that visual attention is supposed to be driven by low-level stimulus in the scene, such as intensity, color, and orientation. The bottom-up methods mainly consist of the following two steps: low-level feature extraction from multiple scales and saliency computation. The saliency is usually computed by center-surround operation using multiple features. After normalization and linear/non-linear combination, a saliency map is computed to represent the saliency of each pixel. On the other hand, the top-down methods are task-driven, where human observation behaviors are exploited in top-down methods to accomplish specific aims. Several studies consider the visual attention as a problem of salient object detection where most bottom-up methods with low-level features

do not have the notation of objects. The results generated by top-down methods usually can provide more semantic and complete user attention information [4].

Focus of attention on distinct patch, the core idea is described as follows: if an image is divided into a number of patches and observed that some distinct patches have some prominent features, such as color, shape, texture, etc., and the numbers are relatively small, while the vast majority of patches' features are generally similar to others. Because the local perspective to consider pixels that have highlight characteristics, and the global view to calculate the distance between salient patches and common patches. As shown in Figure 1, a plane in the sky, it occupies less proportion in the whole picture, but it is the focus of visual attention. Similarly, in the video sequence images, the significantly moving object will receive more attention than surroundings on human visual attention.

Fig.1. Image into number of patches

The single image will be divided into n number of patches they are said to be image patches the same process will also be carried out in the video the video will divided into n number of frames and it is spitted according to the frame size and then the frame will be converted into the gray scale image then from the gray scale image it will forwarded to the pixel segmentation then the pixel will leads to the boundary replacement process [5].

Many methodologies have been proposed to detect objects in an image or video. Instead of the traditional center-surrounding weight mechanism, more and more methods employ the concept “background prior”, which can be regarded as a set of cues or templates for the background and shows a higher performance against conventional salient object detection methodologies However, this is not always the case: in certain circumstances, the image boundary is vulnerable for background priors following two characteristics:

1) *As accurate as possible*. In other words, background priors should contain minimum or even no contents of the foreground.

2) *As sufficient as possible*. That means background priors ought to include as much background as we can.

The most relevant field to our topic is video object segmentation via background subtraction. Different from background subtraction, background priors don’t need to cover all background regions [6].

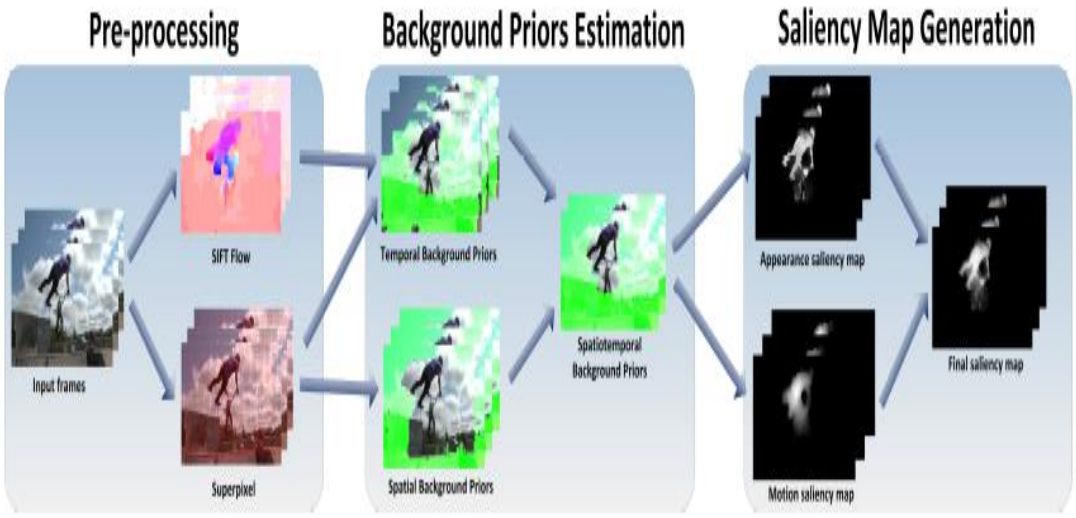

Fig.2. Framework for Saliency map generation

2. Objectives

a. To analyze the accuracy when salient object are located near the boundary region.

b. To improve the RGB resolution when the foreground and the background acquire the same color means it is difficult to adopt.

c. To analyze the Mean Absolute Error (MAE) for every frame in the low resolution.

d. To enhance the robustness of the image.
e. To evaluate the precision of an image/frame by implementing in MATLAB for better performance.

3. Review of literature

Papazoglou, et al, (2013) discussed about the unsupervised video segmentation by tracking many figure-ground segments, it would be too hopeful to presume that a perfect segment is always present in the opening segment pool. Besides, the greedy matching procedure it can create an suboptimal results and it simultaneously tracks multiple holistic figures in the video segment and it perform constantly well on other video and it perform the efficient result. Long-term appearance models are learnt by using a regression-to-overlap framework on many segments to track the initialized segment proposals in this pool. They have derived some samples of training examples for many segment tracks; we are able to track hundreds of segments efficiently by exploiting structures in least squares regression. It ought to improve the appearance and temporal consistency in the video sequences, and also they can perform by taking different size of video sequences for the improvement of the computation time.

Keren Fu, et al, (2014) presented a unified approach to construct graphs for salient region detection in videos. Spatial-temporal saliency image will be smoothed ("appearance +motion +smoothing"), and they are very useful in the performance improvement under large margin. First they should review about the common way of graph construction for still images. After the review been on the image then the graph construction is done on videos and results shows they are directly employing a previous salient region detection method to the current graph construction. Then they are planned to construct a graph that is integrated with both static and motion cues by using a narrative feature: mean histogram of optical flows (MHOF) that will effectively capture the statistical motion information in each super pixel and compute discrepancy/affinity of vertices in a graph. The advantage of this algorithm is video processing is shown by applying the manifold ranking based method to constructed graphs on seven videos. It captured the statistical motion information in each super pixel. It must

concentrate on the optical flow estimation that is not stable and the color contrast can be improved in the object.

Y. Qin, et al, (2015) suggested that Cellular Automata, an intuitive updating mechanism that is planned to develop the intrinsic connectivity of salient objects through connections with neighboring object. Since, there are many effective methods which deal with saliency detection in this each of them has their own superiorities. A well-organized method is used to combine the global distance matrices and then by applying the Cellular Automata to optimize the prior maps via exploiting local similarity. To be depth in this process, it is intended to construct global color distinction and spatial distance matrix that is based on clustered boundary seeds. They have subjected a context-based propagation that has the highest precision and recall and great efficiency and robustness in optimizing other offered methods. The drawback is that in the runtime, they still can achieve the saliency map and the accuracy.

W. Wang, et al, (2015) discussed about the unsupervised method that take part in geodesic distance into saliency empowered video object segmentation. This method evolves suitably in the locations of an object that lies in the background and also it gains the uniform saliency maps.. Spatio-temporal edge map is exposed to point the location of foreground and background. This work is designed to combine SPATIOTEMPORAL EDGE MAP i.e. *an image indicating where edges are in the image which also uses Spatio-Temporal granuality and* GEODESIC DISTANCE to find a pinpoint spatio-temporal prominent result to object segmentation in a prior manner. This will produce a Spatio-temporal saliency maps via the computation of geodesic distance to the estimated background on the inter-frame graph for each pair of adjacent frames. Finally, we computed the segmentation result by combining saliency, global appearance model and location model into the graph-cut energy minimization. The resulted work effects in the accuracy to Spatio-temporal saliency and it performs well in image segmentation process. The downside will be regarding the resolution of an image, and the pixel quality will be less in the probability.

Lei Huang, et al, (2016) studied about the salient object detection method in sequence video of computer application, one important reason to ensure our good performance is that we explore both spatial and temporal cues to detect the saliency on key frames and improve the saliency coherence with bidirectional propagation. Though there have been plenty of video key frame extraction methods, we simply use uniform sampling for efficiency, i.e., sampling

a key frame every video frames and it fully exploring the possibility of Spatio-Temporal difference and coherence of color contrast in object and they have presented that the salient object can be detected in the complex scenes. And globally normalizing that all the saliency maps should obtain the salient object detection those results in the whole video. There will be a trade-off between the effectiveness and efficiency that determines the interval of key frame selection. Large interval may improve the efficiency, but bring in the risk of generating low quality saliency maps. They ought to improve the high quality of salient maps in video frames, if the video is in the large intervals

Bo Xiao, et al, (2017) obtained a hierarchical Boolean Map approach which area able to detect the saliency in images and video. To achieve real time performances and to avoid the flickers, they developed a video consistency by using a border median filter, it performs better in the performance of most feature contrast based methods and frequency based methods. And our method can deal with those cases when the salient object touches the image border, there utilize the post-processing approach. We also evaluate the speed performance. Comparing this with other image saliency detection methods that perform well on detecting the background properly and to improve speed in high definition salient detection objects. In order to achieve the real time runtime performance, we extend the BMS method to a hierarchical one and implement it on GPU. In turns out this method is 10 times faster than BMS, in particular for high-definition images. Also we cope with the consistency problem through a border median filter. Our algorithm is computational efficient, and achieves decent results both in images and videos. It lacks on incorporating temporal information and maintenance of temporal coherence.

Trung-Nghia Le, et al, (2018) studied the Spatio-Temporal Conditional Random Field(STCRF) that captured temporal consistency of regions over frames in the videos and to find the spatial relationship between regions and it is significantly outperforms the methods on publicly available datasets. They take into account temporal information in a video as much as possible in different ways, namely, feature extraction and saliency computation. Visual saliency is also used for estimating human gaze. For salient object detection, object boundaries should be kept as accurately as possible while for human gaze estimation, they are not. Rather, gaze fixation point should be precisely identified and the area nearby the fixation point had better be blurred to present saliency using a Gaussian kernel, for example.

Applying our method directly to gaze estimation is thus not suitable. However, the idea of combining local and global features will be interesting even to gaze estimation. It also applied our method to the video object segmentation task, showing that our method outperform good compare to the previous unsupervised methods. It should focuses on salient object detection and object boundaries which is accurate for human gaze estimation and should improve the gaze estimation videos from side to side application.

Jia Li, et al,(2018) studied that Salient object detection is a hot topic in the area of computer vision. They are, hundreds of innovative models have been proposed for detecting salient objects in images, which have gradually evolved to use the deep models from bottom-up in the large-scale images in the dataset. However, these issues will be there on all video-based SOD has not been satisfactorily explored due to the lack of large-scale video datasets. The most challenging step in constructing such a dataset is providing a reasonable and unambiguous definition of salient objects from the spatiotemporal perspective. Here the VOS, which is having a large-scale dataset with 200 videos. Different from existing datasets, salient objects in VOS are defined by combining human fixations and manually annotated objects throughout a video. As a result, the definition and annotation of salient objects in videos become less ambiguous. Moreover the saliency-guided stacked auto encoders for video-based SOD, and they are compared with massive state-of-the-art models on VOS to demonstrate the challenges of video-based SOD as well as its differences from and relationship with image, based SOD. We find that VOS is very challenging, as it contains a large number of realistic videos, and its subset VOS-E serves as a good baseline for extending existing image based models to the spatiotemporal domain. Moreover, its subset VOS-N covers many real-world scenarios that can facilitate the development of better algorithms. This dataset can be very helpful in video-based SOD, and the unsupervised saliency-guided stacked auto encoders can be used as a good baseline model for benchmarking new video-based models. The drawback is that the MAE will be low it can be improved and the collection of the dataset is less.

Wenguan Wang, et al, (2018) presented a deep learning method in the case of fast video saliency detection using convolution neural networks. The future deep video saliency model has two modules, namely static saliency network and dynamic saliency network, which are designed for capturing spatial and sequential scenes. The saliency estimates from the static

saliency network is incorporated in the dynamic saliency network, which enables our method to automatically learn the way of fusing static saliency into dynamic saliency detection and directly produce final spatiotemporal saliency results with less computation load. Furthermore, we proposed a novel data augmentation technique for synthesizing video data from still images, which enables our deep saliency model to learn generic spatial and temporal saliency and prevents over fitting. Experimental results have shown that our methods generate high-quality salience maps. Additionally, our model is very efficient with a frame rate of 2fps on a GPU. It ought to improve the runtime and the quality is in the low resolution.

Karthik Muthuswamy, et al, (2015) presented a Spatio-temporal saliency maps and color as cues. The framework has two fronts. Firstly, a low-cost method is used to calculate full-resolution saliency maps based on dominant features. Which show that this measure could be utilized to measure spatial as well as temporal saliency and suggest a new technique to combine the two measures, based on the motion present on each frame, to generate the final Spatio-temporal saliency map. Secondly, it is advice a new technique that utilizes the Spatio-Temporal saliency map to identify and track the most salient object present in a given video. While particle filters have been utilized to perform object tracking, the projected method provides an automated mechanism to initialize, track and self-correct itself to identify the 'most salient' object in the video. The performance is evaluated on segmentation datasets. There also develop a simple algorithm to create a Spatio-Temporal saliency map that could tend to outperforms on the saliency object in many state-of-the-art methods. It employs color for its robustness against structural changes in the object while the Spatio-Temporal saliency measure allows the filter to quick converge. The shortcoming are to intelligently resize the frames of a video for salient object detection framework and compare to the accuracy it resembles low to the earlier studies

Meijun Sun, et al,(2018) discussed about the robust deep model for the detection of video eye fixations. By studying the mechanisms of human visual attention and memory, it simulates the process of viewing video sequences by human beings, and added both memory and motion information to enable the model to capture the salient points across neighboring frames. With this process, both the previous detection and the motion information were taken into account to achieve the maximum probability of eye fixations. Thorough the experiments

results validate that our planned model among the superiority in comparison with 11 delegate existing state-of-the art methods. Finally, we highlight our main contributions as follows. 1) A deep model for video saliency detection without the need of any preprocessing operations. 2) The memory information was exploited to enhance the model generalization by considering the fact that changes between two adjacent frames inside a video are limited within a definite range, and hence the corresponding eye fixations should remain correlated. 3) Extensive experiments were carried out and comparative results were reported, which not only supported that our projected model is superior in comparison with the previous methods but also validated the robustness of our planned approaches. The drawback is that can be identified to focus more on human brain activities and explore in detail and human memory can be more accurate and robust detections of eye fixation points as well as their saliencies.

Xiaofei Zhou, et al,(2018) presented a novel framework to improve saliency detection results generated by existing video saliency models. The framework consists of three key steps including localized estimation, spatiotemporal refinement, and saliency update. Firstly, there have been considering some of the temporal consistency and strong correlation among temporally adjacent frames, the local temporal window based opinion models is been designed, i.e. localized estimation models, are learned to obtain the temporary saliency map. Such temporary saliency map can preserve the global shape of salient object in a video. Secondly, the appearance incorporates with the motion information simultaneously, spatiotemporal refinement step is deployed to further improve the temporary saliency map and as there are intended to conclude that the saliency map will perform in distinct boundaries. Finally, the final saliency map is used to update the initial saliency map of current frame, which provide more reliable information for processing the next frame. Extensive experiment prefers 4 challenging video datasets, and the result show the best performances in video. The drawback is that effectiveness is less and the precision is not obtained correctly**.**

Qiong Wang, et al, (2017) presented a novel temporal saliency detection method that extracts temporal saliency information from optical flow field sufficiently. This goal is achieved by using the boundary connectivity cue to extract the boundary movement as the global motion. By employing the guided filter, which preserves object edges, before FMBD process, the method can effectively alleviate the global motion effect and pull out the salient

object extracts high in the background. The proposed method is not only used in the situation that only the salient object moves, but it is also robust to complex scenes such as large camera motion and dynamic background. For the evaluation of video salient object detection, two popular datasets, Seg-Track v2 and Fukuchi, are used. Experimental results reveal that the projected method achieves lower MSR and higher AVC scores than the 5 states-of-the-art referenced methods. Since, the future method generates the temporal saliency. They ought to improve performance by combining it with a spatial saliency detection method and to detect the complex video scenes.

Wenliang Qiu, et al, (2018) studied the scarcity of pixel wise video saliency annotation; we try to take advantage of the more accessible resource, such as eye fixations collected by eye tracker. They leverage the eye fixation data to detect the salient object(s) through total variation-based pair wise interaction. Firstly, extended visual seed regions are extracted from the raw eye fixation data through the geodesic distance based seeds mapping. Then, the spatiotemporal relations between salient objects and background are captured by total variation-based pair wise interaction, which can effectively pop out the saliency region and restrain the background noisy. The saliency maps generated by the proposed TVPI have homogenous foreground regions with high salient values, and the background regions are pure with less noisy. Both of the aforementioned characters indicate that saliency maps of TVPI can be segmented more easily and have greater potential to be applied to subsequent tasks. The disadvantage is that the background noisy is not eliminated correctly and the scalability must also improve.

Wei Li, et al,(2017) discuss about the segmentation of moving vehicles from traffic videos by image processing techniques, the up-to-date spatiotemporal saliency object detection is utilized. The main contribution of our work is that, firstly, a new set of spatial feature and temporal feature are extracted, so that moving objects can be segmented easier, secondly, an improved spatiotemporal consistency model is developed. Compared to the current model, our approach has much less complexity. The spatiotemporal saliency map is shown as a gray level image. For moving objects detection, we need to obtain a binary mask using saliency map. To set a proper threshold, we test on different videos and carried out hundreds time of experiment. Unfortunately, we could not get a proper setting for all the videos. But for each video, we set the best threshold to get foreground mask which is used to segment moving

vehicles. Must improve the parameters and thresholds used in this approach could be set more accurate values, and we will compare this approach with more other methods to prove the advantage and continue to improve the approach.

Nitin Kumar, et al, (2017) studied and analyzed training free popular salient object detection methods in presence of Gaussian, Salt & Pepper and Speckle noise. Extensive experiments were performed on two datasets viz. MSRA5K and DUT OMRON each containing 5000 or more images. The performance of this methods is been evaluated in terms of Precision, Recall and F-Measure. It is observed that Context Aware Saliency Detection method gives best Precision on both datasets in presence of any of the three types of noises. Itti's method gives least Precision on both the datasets because of its low resolution saliency maps. Saliency detection using maximum symmetric surround (AM) method that will performs worst in terms of Recall and F-Measure and this issue must be overcome and the SOD method can perform superior even in the occurrence of noise.

Duan-Yu Chen, et al, (2013) Presented to detect the co-salient video objects efficiently and maintaining the correctness, initially there are using the pre attentive system based on KL-divergence. In addition, to update pre attentive patch set for co-salient objects, sparse coding is used for dictionary learning and further discrimination among co-salient objects. Finally there improve in the correctness of the matching across all video frames based on filtering scheme that are been designed. The tested results show that the projected co-salient video objects model has achieved high precision value and reveals its robustness and feasibility in video. It lacks on co-salient areas in the background are not detected properly and it degrades in the efficient and performance.

F. Perazzi, et al, (2012) presented a Saliency Filters for computation of the saliency images, which are based upon the image abstraction into elements and contrast-based; that formulated as high dimensional Gaussian filters. The graph structure facilitates by smoothing the salient objects and significantly improves the performance of this algorithm. In this the performance and the measuring of the precision and recall rate will be estimated in the high recall can be achieved at the expense of reducing the precision and vice-versa so it is important to evaluate both measures together. We perform two different experiments. In both cases we generate a binary saliency map based on some saliency threshold and it lacks on, color contrast may not always be feasible, e.g., In the case of lighting variations, or when foreground and

background colors are very similar and also can get better sophisticated techniques for image abstraction.

Ruxandra Tapu, et al, (2012) an automatic salient object extraction system based on a Spatio-Temporal attention detection framework. The spatial model is developed starting from the region based contrast applied now on video streams, while the temporal model rely on the interest points correspondence and geometric transforms between key frames. Preceding to detecting the key-frames, first there apply a shot detection procedure, based on this graph partition model there combines with a non-linear scale-space filtering mechanism, the method achieve in the performances of high precision (superior to 90%) and recall (superior to 95%) rates. whatever the movie quality and genre and for both abrupt and gradual transitions The technique is robust to complex background distracting motions and does not require any initial knowledge about the object size or shape. Various tests were performed on numerous video sequences to demonstrate the technique effectiveness. Acceptable results were been generated. An important remark needs to be made: for videos with rich texture or presenting a set of objects, the result of a spatial attention model is, in most of the cases, unrepresentative. But, after incorporating the information gathered from the temporal attention model the system is able to detect were interesting action happens. The drawback is that it should develop a much more powerful object segmentation strategy. And it must consider bottom-up information, it is also necessary to add top down information in order to develop a human vision system.

Daniel Cullen, et al,(2012) Discussed about the multiple video processing algorithms which create a flexible, and robust coastal surveillance system that are useful for marine biologists, environmental agents, etc. There tested this approach broadly by using a real coastal video sequences and showed that the system can decrease the length of typical videos up to about 20 times without losing any salient events. The system can noticeably reduce the human operator which involves in the events of interest. Currently, this system does not operate in real time but with a careful implementation on a multi core architecture real-time performance is within reach. In the course of research, there have made a few observations. Although by fixing the dictionary for each class of objects derived from Google Images has worked quite well, there managed to get better results by manually creating each dictionary from videos in the dataset (different from the video under analysis but captured by the same

camera). This is due to the similarity of viewpoint, distance, and background between all videos. Which are more useful in practice but would require operator involvement to populate/prune the dictionaries? And then they have observed that the detection step works better if objects with significantly different shape are treated as a separate class. For example, the sailboats will be in a large triangular shape that is quite dissimilar from the hull/cabin of motorboats; the detection will works as a best in this case. If there are using to separate the dictionary for each type of boat. Finally, a more advanced classification method than the nearest-neighbor (NN) classifier tested here can be applied. The runtime must be improved this case and the contrast in the image can be better in the resolution.

4. Methodology

4.1 The study prevails that there are many issues regarding the salient object detection in the video stream. The fundamental study is to design a robust salient object detection algorithm to rectify these issues such as accuracy and precision with better resulted foreground image from the background priors. From this time forth, we shall make an effort to model a Water Flow Driven method using Minimum Barrier Distance (MBD) for Salient object detection in this study. The MBD is stronger in the pixel based value and consumes the low computational cost. In this the Water flow from source to the destination pixel with the flow cost. The flow cost is been noticeable only by this MBD. By this approach, the required output of the image with better performance in precision and MAE is achieved.

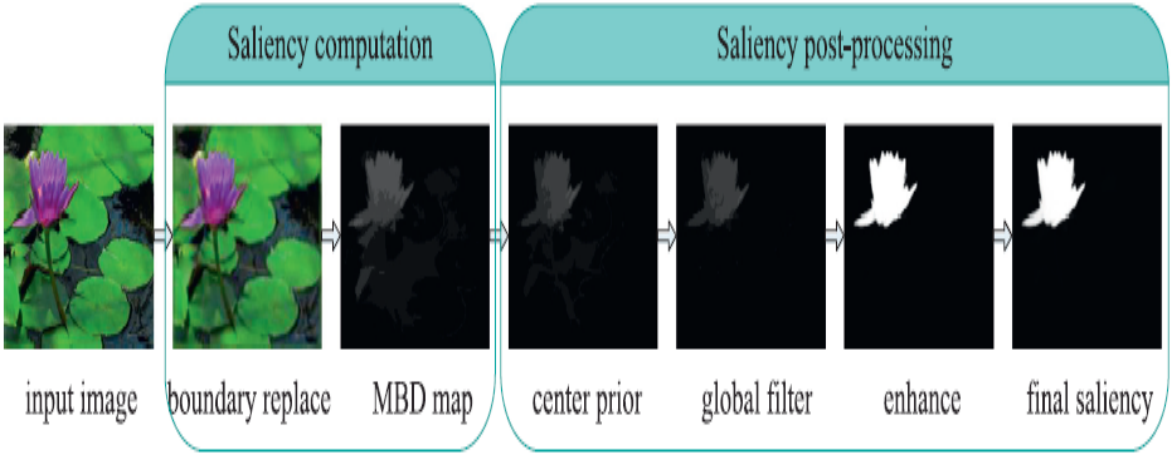

Fig.3. Basic work in Water flow driven method

The workflow of the this proposed method will be shown on the fig.3 here the input video will be splittd into the frames and the input frame will be convereted into the grayscale image it will be binary image of 0 and 1 and then the image segmentatin then the boundary will be replaced by the segmentaion process and the pixel based will be done by the MBD then the center will be prior and obtained the foreground mask then the smoothing will be done on thesalient post processing computation then the RGB frame/image will be obtained and it will be satisfy all the issues that are disccused above.

4.2 Proposed Work Flow

Proposed work flow for the research is as follows:

a. From the input image, the gray scale image is been converted.

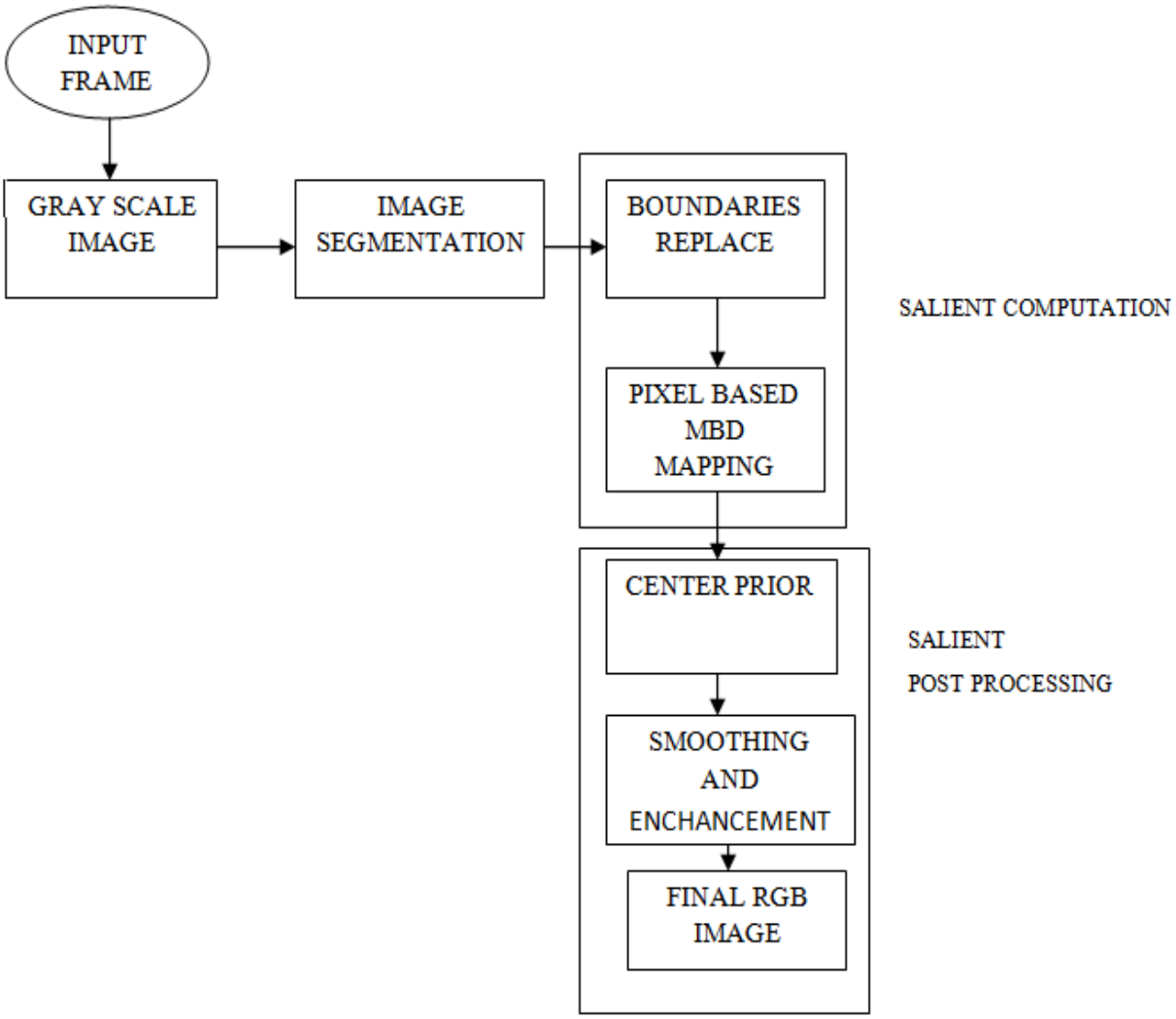

Fig.4. Proposed Workflow

b. In the extraction process, the image is segmented into pixel.
c. Then the boundary is been derived from the pixel and then the MBD will map the salient part.
d. The water will flow until the whole image is flooded for three times until all the pixels are covered.
e. Here, the foreground and the background will be detected and the foreground is been focused for the further process, as center priors and thus the replication of the color will be avoided.
f. Then the precision and the MAE will be calculated by smoothing and enhancement of an image to obtain the required result.
g. Finally, the resulted output with high resolution including the robustness will be achieved with performance speed.

5. Significance of Research

a. The salient object detection mainly focuses on the foreground detection from the background image. And the superior water flow approach is used to extract the RGB image in high resolution and the MAE in the each frame of the foreground will be improved and the speed performance will be high this is achieved by the fast computation of MBD. The significance work here is the video surveillance in the ATM to detect the any type of potential criminal activities that might be arising in the ATM and to find the Theft by using the water flow approach, the image will be saliently computed by the boundary replacement i.e., The person in the ATM will be detected and the MBD mapping is been done.
b. The water will flow until the whole image is flooded for three times until all the pixels are covered. Here, the foreground and the background will be detected and the foreground is been focused for the further process. And in the Post processing the center priors are identified thus the replication of the color will be avoided. Then in the salient process the image will be enhanced to the smoothing and the final RGB image (particular person) in the ATM will be retrieved in the high performance speed when compare to the previous methods.

6. References

1. Tao Xiy, Wei Zhao, Han Wang, Weisi Lin. (2017) Salient object detection with spatiotemporal background priors for video, IEEE Transactions on Image Processing, 26(7) : 3425-3436.
2. F. Perazzi, Philipp Krähenbühl, Yael Pritch,Alexander Hornung.(2012) Saliency filters: Contrast based filtering for salient region detection, IEEE Conference on Computer Vision and Pattern Recognition 733 – 740.
3. C. Yang, Chuan Yang; Lihe Zhang; Huchuan Lu; Xiang Ruan; Ming-Hsuan Yang. (2013) Saliency detection via graph-based manifold ranking, IEEE Conf. on Computer Vision and Pattern Recognition, IEEE Conference on Computer Vision and Pattern Recognition, 3166 – 3173.
4. Yu-Jie Fu,Guan-Lin Wu,Shao-Yi Chien.(2013) Real-time Salient Object Detection Engine for High Definition Videos, 2013 International Symposium onVLSI Design, Automation, and Test (VLSI-DAT) ,1-4.
5. Yanxi Zhao, Zhenhong Shang, Hui Liu. (2015) Highlight Moving Object Detection Based on Spatiotemporal Saliency in Dynamic Background, Third world conference on complex system (WCCS), 1-5.
6. Tao Xiy, Wei Zhaoy, Han Wang, (2017)Salient object detection with spatiotemporal background priors for video, IEEE Transcation on Image Processing,26(7):3425-34236.
7. Fuxin Li, Taeyoung Kim, Ahmad Humayun, David Tsai, James M. Rehg. (2013) Video segmentation by tracking many figure-ground segments, IEEE International Conference on Computer Vision, 2192 – 2199.
8. Keren Fu, Irene Y. H. Gu, Yixiao Yun, Chen Gong, Jie Yang. (2014) Graph Construction For Salient Object Detection in Videos, 22nd International Conference on Pattern Recognition, 2371-2376.
9. Yao Qin, Huchuan Lu, Yiqun Xu, He Wang. (2015) Saliency detection via cellular Automata, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 110 – 119.
10. Wenguan Wang, Jianbing Shen, Fatih Porikli. (2015) Saliency-aware geodesic video object segmentation, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3395 – 3402.

11. Wenguan Wang, Jianbing Shen, Ling Shao.(2018) Video Salient Object Detection via Fully Convolutional Networks, IEEE Transactions on Image Processing,27(1):38-49.
12. Lei Huang, Bin Luo. (2016) Salient Object Detection via Video Spatio-temporal Difference and Coherence, 12th International Conference on Computational Intelligence and Security (CIS), 218 – 222.
13. Bo Xiao, Bin Wang.(2017) Efficient HD Video and Image Salient Object Detection with Hierarchical Boolean Map Approach, 2017 2nd International Conference on Image, Vision and Computing (ICIVC), 1 – 7.
14. Trung-Nghia Le, Akihiro Sugimoto. (2018) Video Salient Object Detection Using Spatiotemporal Deep Features, IEEE Transactions on Image Processing,27(10):5002-5015.
15. Jia Li, Changqun Xia, Xiaowu Chen.(2018) A Benchmark Dataset and Saliency-Guided Stacked Autoencoders for Video-Based Salient Object Detection,IEEE Transactions on Image Processing,27(1):349-36.
16. Karthik Muthuswamy, Deepu Rajan.(2015) Particle filter framework for salient object detection in videos, IET Computer Vision, 9(3):428-438.
17. Meijun Sun, Ziqi Zhou, Qinghua Hu, Zheng Wang, Jianmin Jiang.(2018) SG-FCN: A Motion and Memory-Based Deep Learning Model for Video Saliency Detection, IEEE Transactions on Cybernetics,1-12.
18. Xiaofei Zhou, Zhi Liu, Chen Gong,Wei Liu.(2018) Improving Video Saliency Detection via Localized Estimation and Spatiotemporal Refinement,IEEE Transactions on Multimedia, 1 – 1.
19. Qiong Wang; Lu Zhang; Kidiyo Kpalma.(2017) Fast Filtering-Based Temporal Saliency Detection,IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 232 – 237.
20. Wei Li, Dhoni Putra Setiawan, Hua-An Zhao.(2017) Spatiotemporal Saliency Detection in Traffic Surveillance, International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC),139 – 142.
21. Nitin Kumar, Maheep Singh, Surendra Singh, Abhimanyu Kumar.(2017) A Study of Training Free Salient Object Detection Methods in Presence of Noise, IEEE International Conference on Signal and Image Processing Applications (ICSIPA),362 – 367.

22. Duan-Yu Chen, Chuan-Yu Lin, Nien-Tzu Yang, Jen-Yu Yu.(2013) Sparse Coding-Based Co-Salient Object Detectionwith Application To Video Abstraction,2013 International Conference on Machine Learning and Cybernetics,03:1474-1479.
23. Federico Perazzi, Philipp Krähenbühl, Yael Pritch, Alexander Hornung.(2012) Saliency Filter : Contrast based filtering for Salient region detection, IEEE Conference on ComputerVision and Pattern Recognition, 733 – 740.
24. Ruxandra Tapu, Bogdan Mocanu, Ermina Tapu.(2012) Salient Object Detection in Video Streams,10th International Symposium on Electronics and Telecommunications,275 – 278.
25. Daniel Cullen, Janusz Konrad, T. D. C. Little.(2012) Detection and Summarization of Salient events in costal Environment, IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance,7 – 1.